DigitalTwin-Agri: A Hybrid Digital Twin Framework for Next-Generation Smart Agriculture

Prabhash Nandan (RA2311003011173)
Dept. of Computing Technologies
School of Computing
SRMIST, Kattankulathur, 603203
prabhash08nandan@gmail.com

Satyam Kumar (RA2311003011174)
Dept. of Computing Technologies
School of Computing
SRMIST, Kattankulathur, 603203
satyamchy2121@gmail.com

Dr. Shakeela Sathish (Professor)
Dept. of Computing Technologies
School of Computing
SRMIST, Kattankulathur, 603203
shakeels1@srmist.edu.in

***Abstract*—Modern agriculture increasingly relies on data-driven systems to address challenges such as climate uncertainty, inefficient resource usage, and variability in crop productivity. This study proposes a hybrid Digital Twin framework that integrates sensor-based environmental monitoring, process-oriented crop growth modeling, and machine learning techniques to enable intelligent agricultural decision-making. The framework combines key environmental variables, including temperature, rainfall, humidity, soil moisture, and solar radiation, with crop growth indicators such as leaf area index and biomass to simulate crop behavior and predict yield outcomes. To enhance model robustness under limited data availability, a physics-guided synthetic data generation approach is incorporated. In addition, a feedback-driven updating mechanism continuously refines model parameters based on prediction discrepancies, improving system adaptability over time. Experimental evaluation demonstrates that the proposed hybrid approach enhances predictive accuracy and supports efficient resource management. The results highlight the potential of Digital Twin technology in developing scalable, adaptive, and sustainable smart agriculture systems.**

***Index Terms*—Digital Twin, Smart Agriculture, Internet of Things (IoT), Crop Monitoring, Yield Prediction, Machine Learning, Precision Agriculture, Data-Driven Farming**

I. Introduction

Agriculture remains a fundamental pillar for global food security, yet it is increasingly affected by challenges such as climate variability, water scarcity, soil degradation, and the rising demand for higher productivity [1]. Traditional farming practices, along with early precision agriculture techniques, primarily depend on periodic observations and static decision-making strategies. Such approaches often fail to respond effectively to rapidly changing environmental conditions, leading to suboptimal resource utilization and reduced efficiency [2].

Recent advancements in Internet of Things (IoT), cloud computing, and artificial intelligence have enabled the development of smart agriculture systems capable of continuous data collection and real-time monitoring [2]. While these systems provide valuable insights, most of them function as passive monitoring platforms, offering descriptive analysis rather than predictive and adaptive decision support. As a result, their ability to assist farmers in proactive planning and scenario evaluation remains limited [3].

To overcome these limitations, Digital Twin technology has gained attention as a powerful paradigm for modeling and simulating physical agricultural systems in a virtual environment [3]. A Digital Twin creates a dynamic and continuously updated representation of real-world processes by integrating sensor data, environmental conditions, and computational models. In agricultural applications, this enables simulation of crop growth, detection of stress conditions, and evaluation of management strategies such as irrigation and fertilization before actual deployment [4], [5].

However, existing Digital Twin implementations in agriculture often face significant challenges. Many approaches lack adaptive feedback mechanisms, making it difficult to update models based on real-world deviations. Additionally, integrating heterogeneous data sources and accurately capturing both biological processes and complex data-driven relationships remain open problems [3]. Purely process-based models provide interpretability but lack flexibility, whereas purely data-driven models may achieve high accuracy but often suffer from limited explainability.

In this context, this work proposes a hybrid Digital Twin framework that combines process-based crop growth modeling with machine learning techniques and a feedback-driven adaptation mechanism. The proposed system integrates environmental parameters such as temperature, rainfall, humidity, soil moisture, and solar radiation with crop growth indicators including leaf area index and biomass to improve yield prediction and crop health assessment.

A key contribution of this work lies in the incorporation of a closed-loop feedback mechanism that continuously refines model parameters, enabling improved adaptability and long-term accuracy. The proposed framework aims to provide a scalable and intelligent solution for next-generation agriculture by supporting predictive analytics, informed decision-making, and efficient resource management, aligned with the vision of Agriculture 4.0.

II. Related Work

Digital Twin technology has gained increasing attention in the field of smart agriculture due to its potential to enhance resource management and decision-making processes. Several studies have explored the application of Digital Twins for irrigation management by integrating real-time sensor data with crop growth models. These approaches have demonstrated improvements in water efficiency while maintaining crop productivity; however, most implementations are limited to specific crops or controlled experimental conditions, restricting their applicability in large-scale open-field environments [6].

Soil moisture prediction is another area where Digital Twin-based approaches have been actively investigated. Existing methods combine physically grounded simulations with machine learning techniques to achieve higher prediction accuracy. While these hybrid approaches provide better representation of soil dynamics, they often require significant computational resources and may face challenges in scalability when applied to extensive agricultural regions [7].

In greenhouse environments, Digital Twin frameworks have been utilized for microclimate monitoring and sensor optimization. Reinforcement learning-based strategies have been proposed to dynamically select sensors and improve monitoring efficiency. Despite these advancements, such systems are typically designed for controlled indoor settings and do not fully address the complexities associated with open-field agriculture and large-scale crop production [8].

Furthermore, Digital Twin architectures have been applied in hydroponic and robotic farming systems to automate agricultural operations. These systems leverage artificial intelligence and simulation-driven control mechanisms to enhance operational efficiency. However, their dependence on controlled environments and specialized infrastructure limits their adaptability to conventional farming practices [9].

Recent review studies indicate that many existing Digital Twin implementations in agriculture remain at an early stage, often functioning as digital shadows that primarily focus on monitoring and visualization rather than predictive modeling and adaptive control [10]. Key challenges identified include lack of standardization, interoperability issues, and difficulties in integrating heterogeneous data sources across different agricultural systems.

Additionally, Digital Twin approaches have been explored for predictive maintenance of agricultural machinery using sensor data and deep learning models. While these methods improve equipment reliability and fault detection, they require detailed system knowledge and high-fidelity modeling, which may not always be feasible in practical agricultural settings [11].

Overall, the existing body of research highlights the significant potential of Digital Twin technology in agriculture while also revealing critical limitations, including scalability constraints, lack of adaptive feedback mechanisms, and insufficient integration of physical models with data-driven approaches. These limitations motivate the development of a hybrid Digital Twin framework that combines process-based modeling, machine learning, and feedback-driven adaptation, as proposed in this study.

III. Dataset Description

The dataset used in this study is developed by combining agricultural production statistics with satellite-derived agroclimatic information to support data-driven modeling within the proposed Digital Twin framework. The primary agricultural data are obtained from the FAOSTAT database, which provides standardized and reliable records of crop production. The dataset includes annual observations for India from 2000 to 2024, covering key variables such as harvested area (in hectares), crop yield (in kilograms per hectare), and total production (in tonnes). These variables serve as reference indicators for training and evaluating the predictive models.

To incorporate environmental influences, climatic data are sourced from the NASA POWER dataset, which provides satellite-based measurements of agroclimatic conditions. The selected parameters include mean air temperature, relative humidity, total precipitation, and surface solar radiation. These variables capture the environmental conditions that significantly influence crop growth and yield. The climatic data are aggregated at an annual level to ensure consistency with the agricultural dataset.

Prior to integration, a structured preprocessing pipeline is applied to ensure data quality and consistency. This includes temporal alignment by selecting a common time range (2000–2024), standardization of variable names, and verification of measurement units. Data cleaning procedures are performed to handle missing values and remove inconsistencies. The agricultural and climatic datasets are then merged using the year as a common reference key, resulting in a unified dataset that links crop performance with environmental conditions.

Despite providing reliable historical information, the dataset remains limited in size due to its annual resolution and data availability constraints. Such limitations can impact the generalization capability of machine learning models. To address this issue, a physics-guided synthetic data generation approach is employed to expand the dataset. Synthetic samples are generated by introducing controlled variations within realistic agronomic ranges, ensuring that the generated data remain consistent with real-world agricultural behavior.

To validate the quality of the synthetic data, a distribution-based comparison is conducted against the original dataset using percentile analysis. This ensures that the generated samples preserve the statistical characteristics of real observations without introducing unrealistic patterns. The final dataset, comprising both real and synthetic data, provides a robust foundation for training machine learning models and enabling accurate Digital Twin simulations for crop yield prediction.

A. System Overview

The architecture of the proposed hybrid Digital Twin framework is presented in Fig. 1. The system is composed of four primary modules: data acquisition and preprocessing, physical crop growth modeling, machine learning-based prediction, and a feedback-driven update mechanism.

Initially, environmental and crop-related data are collected and processed to ensure consistency and reliability. These processed inputs are then simultaneously provided to both the physical model and machine learning components. The outputs generated are continuously compared with reference values, and the resulting prediction error is utilized to adjust model parameters through a feedback loop. This iterative process enables the system to adapt dynamically and improve its performance over time [17].

B. Physical Crop Growth Modeling

The physical crop growth model describes the biological mechanisms underlying crop development and yield formation. Process-based models are widely adopted to maintain consistency between environmental conditions and crop response while preserving agronomic interpretability [18]. The primary state variables considered in this study include leaf area index (LAI), biomass, and final crop yield.

The canopy growth dynamics are represented using a logistic growth formulation:

$$\frac{dLAI(t)}{dt} = r \cdot LAI(t) \cdot \left(1 - \frac{LAI(t)}{LAI_{max}}\right) \cdot f_T \cdot f_{SM} \quad (1)$$

where r denotes the intrinsic growth rate, LAI_{max} represents the maximum achievable leaf area index, and f_T and f_{SM} correspond to temperature and soil moisture stress factors, respectively.

Biomass accumulation is modeled based on radiation interception theory [20]:

$$IPAR(t) = PAR(t)\left(1 - e^{-k \cdot LAI(t)}\right) \quad (2)$$

$$\frac{dB(t)}{dt} = RUE \cdot IPAR(t) \cdot f_T \cdot f_{SM} \quad (3)$$

where $B(t)$ denotes biomass, RUE is the radiation use efficiency, and k represents the light extinction coefficient.

The final crop yield is estimated using the harvest index formulation [21]:

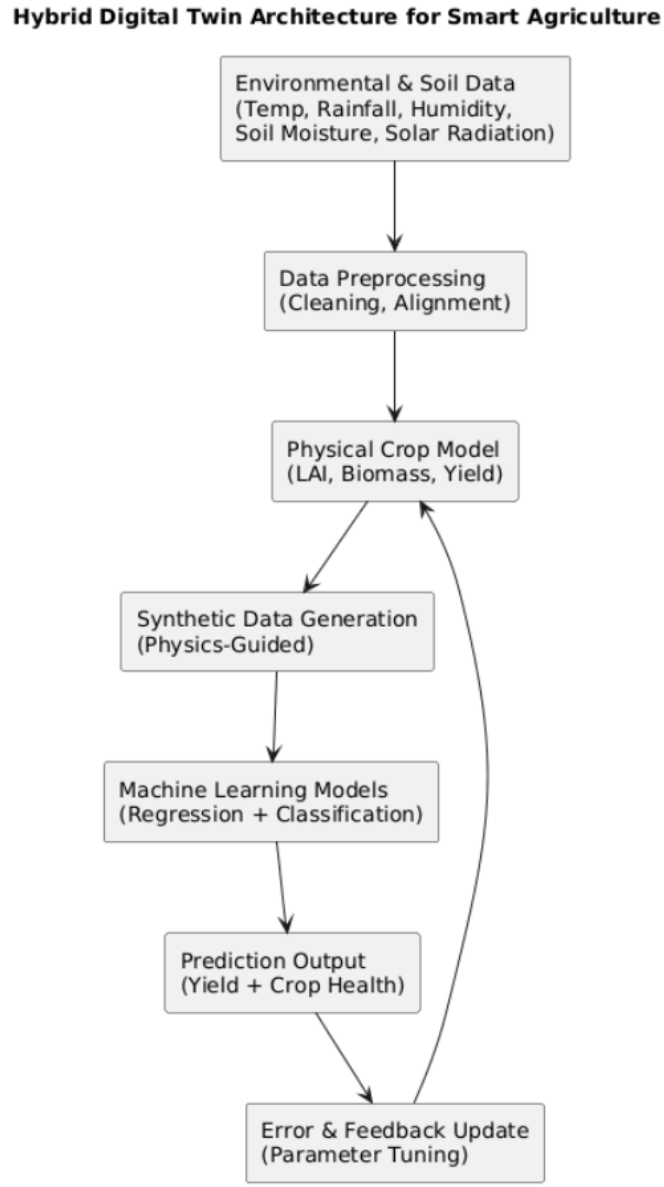

Fig. 1. Architecture of the proposed hybrid Digital Twin framework integrating physical modeling, machine learning, and feedback-based adaptation.

$$Y = HI \cdot B_{final} \quad (4)$$

where HI denotes the proportion of economic yield relative to total biomass.

C. Physics-Guided Synthetic Data Generation

To address the limitation of insufficient large-scale agricultural datasets, a physics-guided synthetic data generation strategy is adopted. Instead of applying arbitrary augmentation, environmental variables are varied within realistic agronomic bounds, and the physical crop model is used to simulate the corresponding crop responses.

This approach ensures that the generated data remain biologically consistent while increasing dataset diversity. As a result, the machine learning models trained on this augmented dataset exhibit improved robustness and generalization capability. Such physics-informed data generation techniques are commonly employed in Digital Twin applications [22].

D. Machine Learning Models

Machine learning models are incorporated into the Digital Twin framework to capture complex and nonlinear relationships between environmental conditions and crop performance. Regression models are utilized for predicting crop yield, whereas classification models are applied for assessing crop health status.

The integration of machine learning with physical modeling enables the system to combine data-driven learning with domain knowledge. This hybrid approach enhances predictive accuracy while preserving interpretability, outperforming standalone modeling techniques [16].

E. Feedback-Based Digital Twin Updating

A feedback-driven updating mechanism is introduced to continuously refine the Digital Twin. The predictions generated by both the physical and machine learning models are compared with observed or reference values, and the resulting error is used to update model parameters and system states.

This closed-loop process allows the Digital Twin to evolve over time, improving its adaptability and maintaining accuracy under varying environmental conditions. Feedback-based updating is a defining characteristic of advanced Digital Twin systems and is essential for long-term reliability [17].

IV. Results and Discussion

This section evaluates the performance of the proposed Digital Twin-based framework and examines the behavior of physical, machine learning, and hybrid models for crop yield prediction and health classification.

A. Actual and Predicted Yield Analysis

The relationship between actual groundnut yield and the predicted output of the Digital Twin is illustrated in Fig. 2. The scatter plot compares observed yield values with model predictions, along with a reference diagonal indicating ideal agreement.

The predicted values are concentrated within a relatively narrow range, whereas the actual yield values exhibit greater dispersion. This suggests that the Digital Twin effectively captures the baseline yield driven by environmental conditions and crop growth dynamics. The observed variability in actual yield can be attributed to external influences such as irrigation

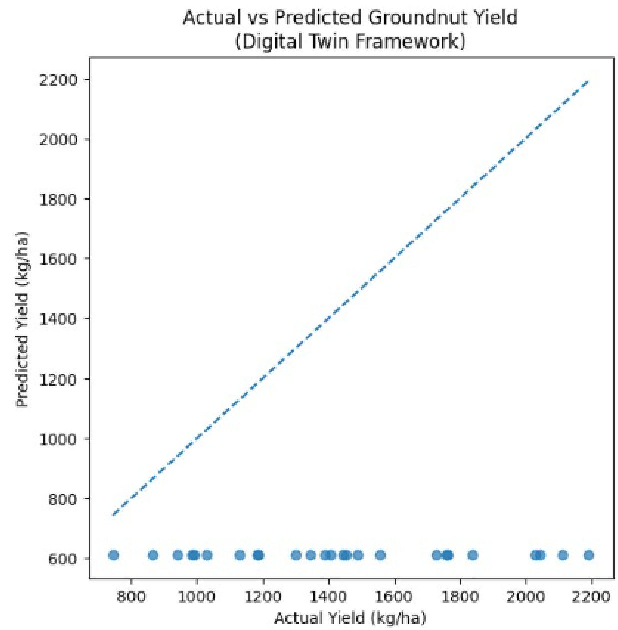

Fig. 2. Real and modeled groundnut yield using the Digital Twin framework.

practices, soil management, and technological improvements that are not fully represented in the model.

B. Temporal Yield Trends

The temporal variation of actual and predicted yield is presented in Fig. 3. The actual yield demonstrates an increasing trend over time with noticeable fluctuations.

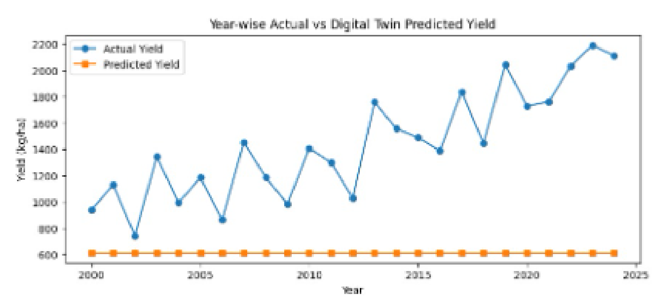

Fig. 3. Comparison of actual and predicted yield over time.

In contrast, the predicted yield remains comparatively stable across the observed period. This indicates that the Digital Twin primarily reflects intrinsic crop behavior under consistent environmental inputs, whereas the upward trend in actual yield is likely influenced by advancements in agricultural practices such as improved seed varieties, fertilizer usage, and mechanization.

C. Crop Health Classification Performance

The performance of crop health classification is evaluated using the confusion matrix shown in Fig. 4. The classification framework categorizes crop conditions into poor, moderate, and good based on the state variables derived from the Digital Twin.

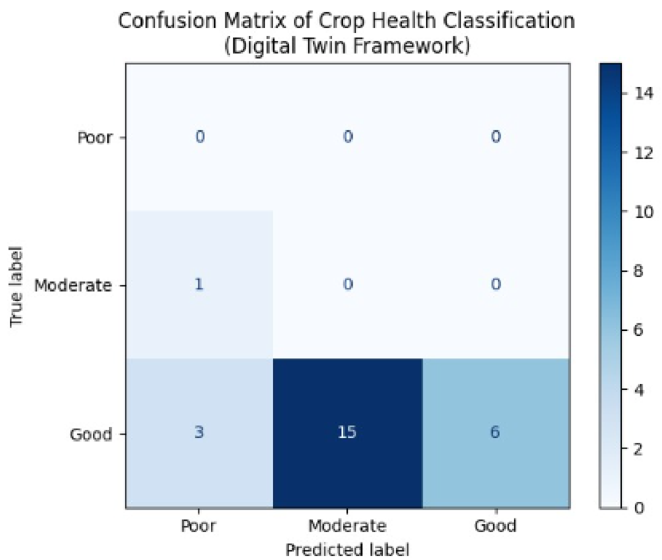

Fig. 4. Confusion matrix for crop health classification.

The results indicate that the model achieves strong classification performance for moderate and good crop conditions. However, the accuracy for poor conditions is relatively lower, which can be attributed to feature overlap and imbalance in class distribution. This highlights the need for additional representative data to improve classification performance in less frequent conditions.

D. Model Performance Comparison

A comparative analysis of prediction error across different models is shown in Fig. 5. The physical model exhibits higher error due to its simplified representation of complex agricultural processes.

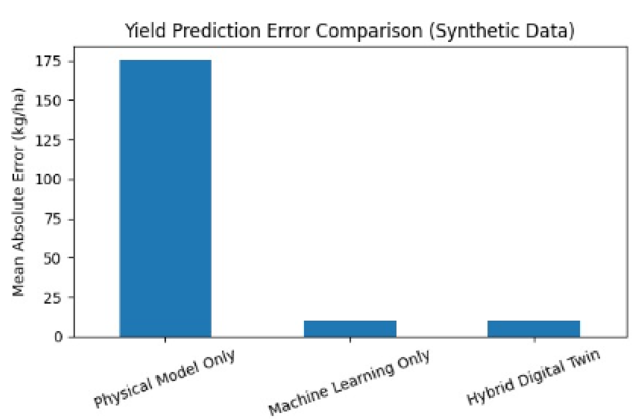

Fig. 5. Comparison of mean absolute error for different models.

The machine learning model demonstrates improved performance by capturing nonlinear relationships present in the data. The hybrid Digital Twin model achieves competitive accuracy by integrating both physical understanding and data-driven learning, thereby balancing interpretability and predictive performance.

E. Sensitivity Analysis

The sensitivity of predicted yield to key environmental variables is analyzed in Fig. 6 and Fig. 7. The temperature response curve in Fig. 6 follows a bell-shaped pattern, indicating that yield increases up to an optimal temperature range and decreases beyond it.

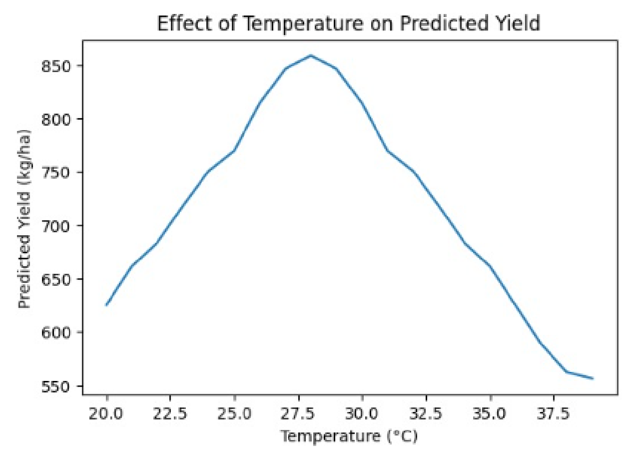

Fig. 6. Effect of temperature on predicted yield.

Similarly, Fig. 7 illustrates the relationship between soil moisture and yield. The results show that yield improves with increasing soil moisture up to an optimal threshold, beyond which excessive moisture negatively impacts crop growth.

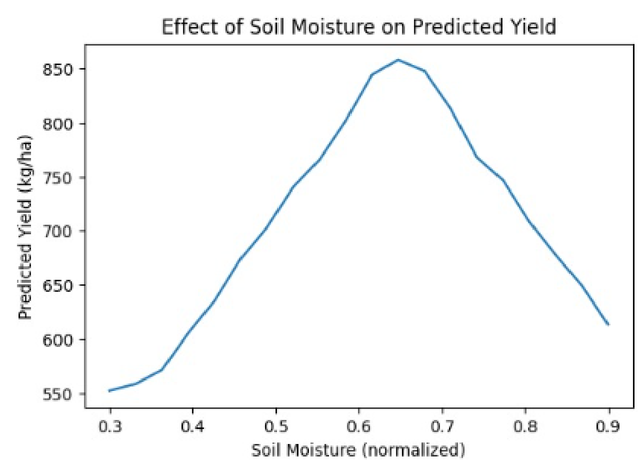

Fig. 7. Effect of soil moisture on predicted yield.

These patterns are consistent with established agronomic knowledge, indicating that the model captures realistic crop-environment interactions.

F. Quantitative Performance Metrics

The quantitative evaluation of model performance is summarized in Fig. 8, which presents the comparison of prediction metrics.

	Model Type	Algorithm Used	MAE (kg/ha)	RMSE (kg/ha)	R^2
0	Physical Model Only	Deterministic Physics Equations	175.238	40201.628	-1.573
1	Machine Learning Only	Linear Regression	10.361	178.212	0.989
2	Hybrid Digital Twin (Proposed)	Physics + Linear Regression (Residual)	10.361	178.212	0.989

Fig. 8. Performance comparison of yield prediction models (Table I).

The physical model shows comparatively higher prediction error, whereas the machine learning and hybrid Digital Twin models achieve better accuracy. The hybrid approach, in particular, demonstrates improved consistency, highlighting the advantage of combining physical modeling with data-driven techniques.

G. Discussion Summary

Overall, the results demonstrate that the proposed Digital Twin framework effectively models baseline crop yield behavior and responds realistically to environmental variations. The hybrid approach combines the strengths of physical modeling and machine learning, providing reliable predictions and improved decision support for smart agriculture.

V. Conclusion

This paper presented a hybrid Digital Twin framework for smart agriculture by integrating physical crop modeling, machine learning, and a feedback-driven update mechanism. The proposed approach improves yield prediction and crop health assessment by combining domain knowledge with data-driven learning.

Experimental results demonstrate that the hybrid model achieves better predictive performance compared to standalone physical and machine learning approaches. The inclusion of a feedback mechanism further enhances system adaptability by enabling continuous refinement under changing environmental conditions. Overall, the proposed framework offers a scalable and effective solution for supporting intelligent decision-making and efficient resource utilization in modern agriculture.

VI. Future Scope

The proposed Digital Twin framework can be further extended by integrating real-time IoT sensor networks to enable continuous data acquisition and real-time synchronization with physical agricultural environments. This would improve system responsiveness and allow more dynamic decision-making. Future work can also focus on expanding the framework to support multiple crop types and diverse geographical regions, thereby improving its scalability and generalization capability. Incorporating advanced machine learning techniques, such as deep learning models, may further enhance prediction accuracy by capturing complex nonlinear relationships.

Additionally, the integration of remote sensing data and satellite imagery can strengthen environmental monitoring and provide more detailed insights into crop conditions. Deploying the framework on cloud and edge computing platforms can improve scalability and reduce latency, making it more suitable for real-world agricultural applications.

Further improvements may include the incorporation of economic and management factors, such as irrigation strategies, fertilizer usage, and cost considerations, to provide more comprehensive decision support. Enhancing the feedback mechanism with adaptive learning capabilities can also enable the development of a more autonomous and self-evolving Digital Twin system.

Acknowledgment

The authors would like to express their sincere gratitude to SRM Institute of Science and Technology (SRMIST), Kattankulathur, for providing the necessary support and resources to carry out this research work.

The authors also extend their heartfelt appreciation to Dr. Shakeela Sathish for her valuable guidance, continuous support, and insightful suggestions throughout the development of this study. Her expertise and encouragement played a significant role in shaping this work.

Furthermore, the authors acknowledge the use of publicly available datasets from FAOSTAT and NASA POWER, which contributed significantly to the successful completion of this research.

References

[1] "Computers and Electronics in Agriculture," 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666154323002831

[2] "Heliyon Review Article," ScienceDirect, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949736125001332

[3] "A Comprehensive Review of Digital Twins Technology in Agriculture," Agriculture (MDPI), 2025. [Online]. Available: https://www.mdpi.com/2077-0472/15/9/903

[4] "From Mechanistic-Driven to Data-Driven: A Review of Crop Models and Hybrid Approaches," ScienceDirect, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S277237552500930X

[5] "Digital Twin Deployment for Smart Agriculture in Cloud-Fog-Edge Infrastructure," International Journal of Computers and Applications, 2023. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/17445760.2023.2235653

[6] "Use of a Digital Twin for Water Efficient Management in a Processing Tomato Commercial Farm," Agronomy (MDPI), vol. 15, no. 9, 2025. [Online]. Available: https://www.mdpi.com/2073-4395/15/9/2132

[7] "A Digital Twin Approach for Soil Moisture Measurement with Physically Based Rendering Simulations and Machine Learning," Electronics (MDPI), vol. 14, no. 2, 2025. [Online]. Available: https://www.mdpi.com/2079-9292/14/2/395

[8] "A Digital Twin Framework for Sensor Selection and Microclimate Monitoring in Greenhouses," AgriEngineering (MDPI), vol. 7, no. 10, 2025. [Online]. Available: https://www.mdpi.com/2624-7402/7/10/315
[9] "DigiHortiRobot: An AI-Driven Digital Twin Architecture for Hydroponic Greenhouse Horticulture," Future Internet (MDPI), vol. 17, no. 8, 2025. [Online]. Available: https://www.mdpi.com/1999-5903/17/8/347
[10] "Harnessing Digital Twins for Sustainable Agricultural Water Management: A Systematic Review," Applied Sciences (MDPI), vol. 15, no. 8, 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/8/4228
[11] "Digital Twin-Based Approaches for Agricultural Machinery Damage Prediction and Maintenance," Journal of Computational Design and Engineering, vol. 12, no. 10, 2025. [Online]. Available: https://academic.oup.com/jcde/article/12/10/87/8262896
[12] "A Comprehensive Review of Digital Twins Technology in Agriculture," Agriculture (MDPI), vol. 15, no. 9, 2025.
[13] "Application of Digital Twin Technology in Smart Agriculture: A Bibliometric Review," Agriculture (MDPI), vol. 15, no. 17, 2025.
[14] "Digital Twin Deployment for Smart Agriculture in Cloud-Fog-Edge Infrastructure," International Journal of Computers and Applications, 2023.
[15] "Digital Twins in Agriculture: A Review of Recent Progress and Open Issues," Electronics (MDPI), vol. 13, no. 11, 2024.
[16] "Agricultural Digital Twin for Smart Farming: A Review," ScienceDirect, 2025.
[17] "Digital Twin Deployment for Smart Agriculture in Cloud-Fog-Edge Infrastructure," International Journal of Computers and Applications, 2023.
[18] "Digital Twins in Agriculture: A Review of Recent Progress and Open Issues," Electronics, vol. 13, no. 11, 2024.
[19] Food and Agriculture Organization (FAO), "AquaCrop Model." [Online]. Available: https://www.fao.org/aquacrop/en/
[20] "Radiation Interception and Biomass Production," Philosophical Transactions of the Royal Society B, 1977.
[21] "The Biological Yield and Harvest Index of Cereals," Journal of Agricultural Science, 1976.
[22] "Synthetic Data Generation for Agricultural Decision Support Systems," Applied Sciences, 2025.